CENTER FOR THE INFORMATION SCIENCES
LEHIGH UNIVERSITY M, PENNSYLVANIA

EXPERIMENTAL RETRIEVAL
SYSTEMS STUDIES

Report No. 4

Transition to On-Line

Information and Reference Retrieval

by

William A. Smith, Jr.
Ronald R. Anderson
Michael A. Jennings

August 1968

Center for the Information Sciences

Lehigh University

Bethlehem, Pennsylvania 18015

Experimental Retrieval

Systems Studies

Report No. 4

Transition to On-line

Information and Reference Retrieval

by

William A. Smith, Jr.
Ronald R. Anderson
Michael A. Jennings

August 1968

Preface

This report summarizes experience in maintaining an experimental computerized information and reference retrieval system and in its conversion to on-line user interaction by remote access. The Project Director for establishing the document base and the system design was Robert S. Taylor, formerly Director, Center for the Information Sciences, Lehigh University, and currently Director of Libraries, Hampshire College. Primary contributors to the system development and operation were Ronald R. Anderson, Michael A. Jennings, James S. Green and Anthony F. Amico, Research Assistants in the Center for the Information Sciences. Dr. William A. Smith, Jr., formerly Director, Computing Laboratory and now Associate Director for Systems, Center for the Information Sciences, contributed to system design and directed completion of the studies. Experience from this and related studies has been consolidated into further research and implementation of new computing systems, utilizing the established document base.

Previous reports which bear directly on the Office of Naval Research studies are:

Experimental Retrieval Systems Studies, Report No. 2, Systems Manual for the Experimental Literature Collection and Reference Retrieval System. Center for the Information Sciences. R. S. Taylor, R. R. Anderson, A. F. Amico, J. S. Green. April 1967.

Experimental Retrieval Systems Studies, Report No. 3, Part 1, An Associativity Technique for Automatically Optimizing Retrieval Results. Center for the Information Sciences. R. R. Anderson, April 1967.

An Analysis of the Factors of Conversion from a Batch-Processing, Tape-Oriented Retrieval System to an On-Line, Real-Time System. MS Thesis, M. A. Jennings, Lehigh University, 1967.

Table of Contents

List of Figures and Tables

Abstract

An existing experimental document and reference retrieval system operating in batch-processing, tape-oriented mode was converted to an on-line mode with user interaction. Objectives were faster response time, integration of functions, and system accessibility at user location. Effects on file organization, addressing techniques, file maintenance, communication, scheduling and programming are analyzed. Modification of an existing batch oriented system can yield an effective interactive search capability. Realization of full benefits of the on-line concept, particularly maintenance of the document base, requires system redesign rather than extension of existing procedures.

User reaction to the experimental system was favorable although a variety of improvements were suggested for an operational environment. Output of associativity indicators was valuable in reformulating search requests or in adjusting the volume of cited references for both experienced and unitiated users.

I. Introduction

Document reference information storage and retrieval systems have been designed and developed in order to help the user of such a system in his information seeking. These systems in their experimental stages have been generally built on the same basic concepts of design, structure, and system goals. The environment of these systems evolves around a service concept which provides document reference(s) for the user upon his search request to the system. The document reference is any one or a combination of the following:

(a) Document title
(b) Author of the document
(c) Location of the document
(d) Publication date
(e) Issuing agency
(f) Location of the agency
(g) Contract and/or grant numbers
(h) Journal, volume, pages, date
(i) Abstract of the document

To provide the user of such systems with a more easily accessible and expanded service, remote access and machine-aided cognition become experimental tools. With these tools, an intimate collaboration between a human user and a computer has been initiated.

The Reference Retrieval System

The system upon which this study is based is the Experimental Reference Retrieval System of the Center for the Information Sciences, Lehigh University. Two major objectives in maintaining this system are to provide an operating collection for instruction and experimentation and to provide a basic reference system for the papers, books, and reports in the field of the Information Sciences. The system is based on the coordinate index concept.

This experimental document reference and retrieval system includes a collection of over 2500 documents related to the information sciences. Such general subject areas as information retrieval, library automation, management information systems,

artificial intelligence, computer-assisted instruction, logic,
and linguistics are included in the collection. All the documents are housed in either the Center for Information Sciences
(CIS) facilities or the Lehigh University Library.

The selection of documents for inclusion is a cooperative
judgment by the CIS personnel. The document is indexed, using
a coordinate indexing approach, by one or more graduate students
in the Information Sciences program. The index terms used in
describing each document are chosen from a thesaurus of approx-
imately 450 terms. The average is between five and six terms
assigned to each document. The index terms selected, together
with the remainder of the data used for document representation,
are then keypunched on IBM cards and transferred to file storage
by a computer program. Two separate files are maintained,
a serial file and an inverted file.

Two physically separate magnetic tape files are maintained
to support the operational computer programs of this system in
a batch processing, tape oriented (BPTO) mode. These two files
are the serial file and the inverted file.

The serial file is linear or sequential in its structure.
The file is alphabetical by author's last name. Two records,
sixty-four computer words per record, are used for each document
in the serial file. The first record contains author(s),
title, location of document, document number, and the index
term numbers for that document. The maximum number of index
terms per document is twenty-five. The second record contains
bibliographic data for the document. This is the issuing
agency, place, publication data, pagination, journal name,
volume, year, and contract and/or grant numbers. Figure 1
gives an illustration of the fixed field file structure of
the serial file records.

Author	Title	Doc. No.	Term Numbers

Doc. No.	Bibliographic Data

Figure 1.   Serial File Organization

The inverted file is maintained in ascending order by the unique number assigned to each index term. Each record includes the unique term number, the natural language representation of the term, scope notes (appropriate clarification of index terms), a usage frequency count, and the accession numbers of the documents to which the index term is assigned.

Two types of search procedures are available to users of the CIS system, both using term numbers for searching rather than natural language. The basic search program allows the user to formulate a search specification using any combination of from one to four index terms connected by the Boolean operators "and", "or", and "not". Three options are available for print-out of document representations. The second type of procedure is an associative search in which the search specification can be reformulated by inclusion of some of the most highly related terms in the table, and tried again.

Term Number	Document Counter	Term (Alpha)	Scope Note	Document Numbers

.
.
Continuea
.
.

Term Number	Document Counter	Document Numbers

Figure 2. Inverted File Organization

A significant improvement in the system has allowed the conversion of the files to disk storage, allowing a direct access, rather than sequential, form of processing. The file contents and format remain the same. Also, on-line, real time (OLRT) access to the contents of the files is provided by Model 33 Teletypewriter terminals and a General Electric Datanet-15 transmission controller on a dial-up basis to a GE 225 computer. Conversion to this on-line capability from an existing batch oriented operation has been the subject of Office of Naval Research Contract Nonr 610(08), "Studies in On-Line Information and Reference Retrieval". This report discusses the effort required, the benefits achieved and the user reaction based on experience in adapting the experimental CIS system to remote access operation.

The system is completely operational and is described in detail in the form of a systems manual.[1] At the present time,

this and other work is being consolidated under a separate
NSF grant, "Prototype Retrieval System Development within an
Information Resource Laboratory", under the direction of Pro-
fessor D. J. Hillman. The experimental literature collection
of the Center is serving as a data base for the study, which
will utilize improved hardware and software facilities to focus
on further research and experimentation emphasizing on-line
content or fact retrieval.

II. System Characteristics

A. The Components of the Tape-oriented System

The computer system to support the basic document reference and retrieval system can be separated into three categories. These categories, or subsystems, are the search subsystem, the update subsystem, and the listing subsystem. These subsystems were programmed utilizing FORTRAN II with the G.E. 225 computer system. The subsystems are a series of chained programs which were originally developed for a batch processing, tape-oriented (BPTO) operation.

The Search Subsystem

To execute a search on the system, a search statement must be formalized for a particular query using the thesaurus of terms and the Boolean operations. The thesaurus terms must be converted to their respective term numbers and placed with a Boolean statement in a fixed field. For example, a user may wish to search for documents dealing with computer and systems but not programming. His statement would then look like the following:

computers + systems - programming

In the statement, + is the Boolean operator and, V is the or, and - is the not operation. Before this statement can be submitted to the system via a punched card, the respective term numbers must be substituted as in the following example:

(67500 + 88200) - (75500)

The system will accept up to a four-term Boolean search statement.[2] When the search statements are read into the search program, they are equated against the serial file. This search strategy was adopted to increase searching speeds. Instead of searching and performing the Boolean operations on an inverted file, a method of searching only a linear file was devised. A linear file organization for tape-oriented retrieval systems is more efficient than an inverted file organization.[3] To increase searching speed, the serial file records are blocked in groups of eight.

A second type, or associative, search can be called by the user which will display a table of terms and their associative values of co-occurrence with the terms of the query. The associative values are computed by using the inverted file.[4]

This procedure is normally used when one of two cases arises: (1) the searcher wants to learn something about the way in which index terms are used in the system, or (2) the searcher has already tried the basic search program and wants to either increase or decrease the number of documents to be retrieved. The output from this search procedure is designed to measure the relationship between the index terms used in the search specification and the other index terms available in the system.

The resulting document references of a search request are displayed on a printer. The number of searches run at a given time can be "batched" to ten search statements, taking only one pass of the serial file. Figure 3 illustrates the search subsystem in a simplified flow chart.

The Update Subsystem

If a system is designed to be dynamic, update and maintenance operations must be included in the total system. The update operations performed on the files are standard basic operations (Figure 4): adding new documents, deleting old or unwanted documents, adding new index terms, and deleting index terms.

The inputs to these programs are a variety of different punched card formats. At present, 150 new documents can be added during one update run. Deletion can be run the same way with a maximum of fifty-five.

The update programs must run in a two-phase operation due to the limitation of available tape drives. The first phase operates on the serial file only. After the first phase, the tape reels are changed and the inverted file is mounted for the running of the second phase. Obviously, this method of update is time-consuming and consequently costly in terms of computer time charges.

Figure 3. The Search Subsystem

```
┌──────────┐  ┌──────────┐  ┌──────────┐  ┌──────────┐
│ Document │  │ Document │  │   Term   │  │   Term   │
│   Adds   │  │ Deletes  │  │   Adds   │  │ Deletes  │
└──────────┘  └──────────┘  └──────────┘  └──────────┘
```

Key-
punched
Input

Update
Program

Serial
File

In-
verted
File

Serial File
Adds &
Deletes

Inverted
.File
Addﬆ &
Deletes
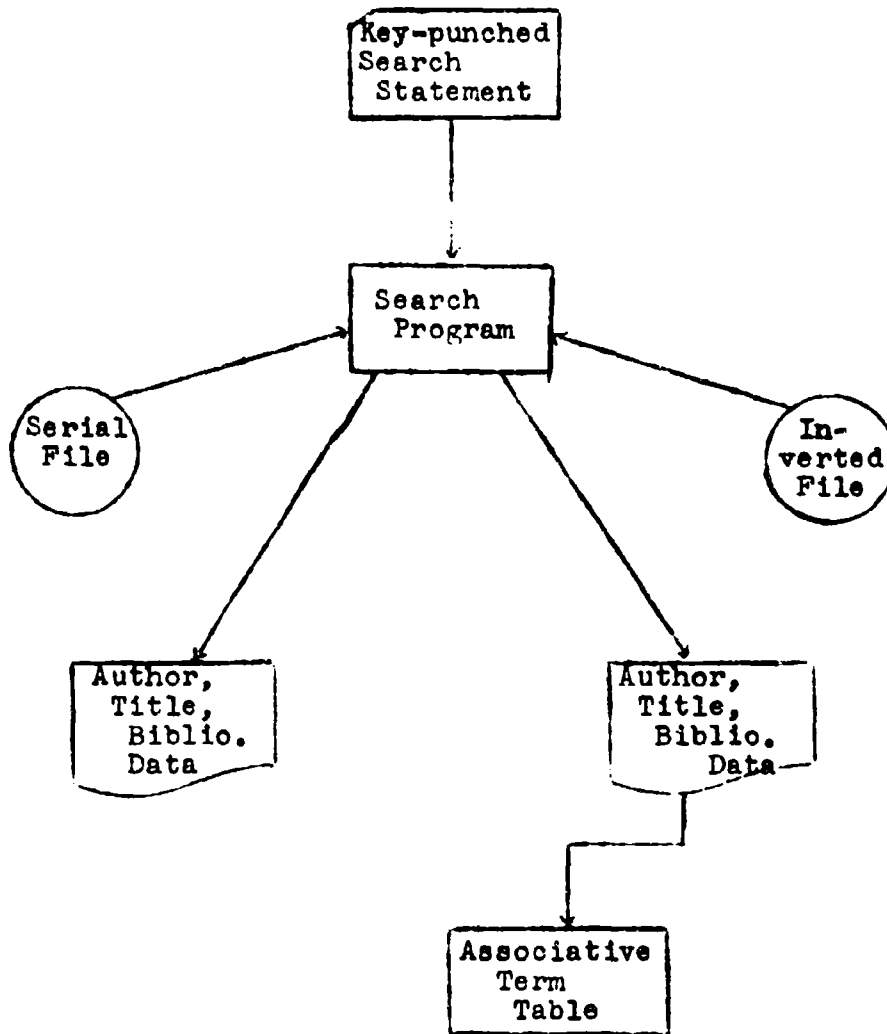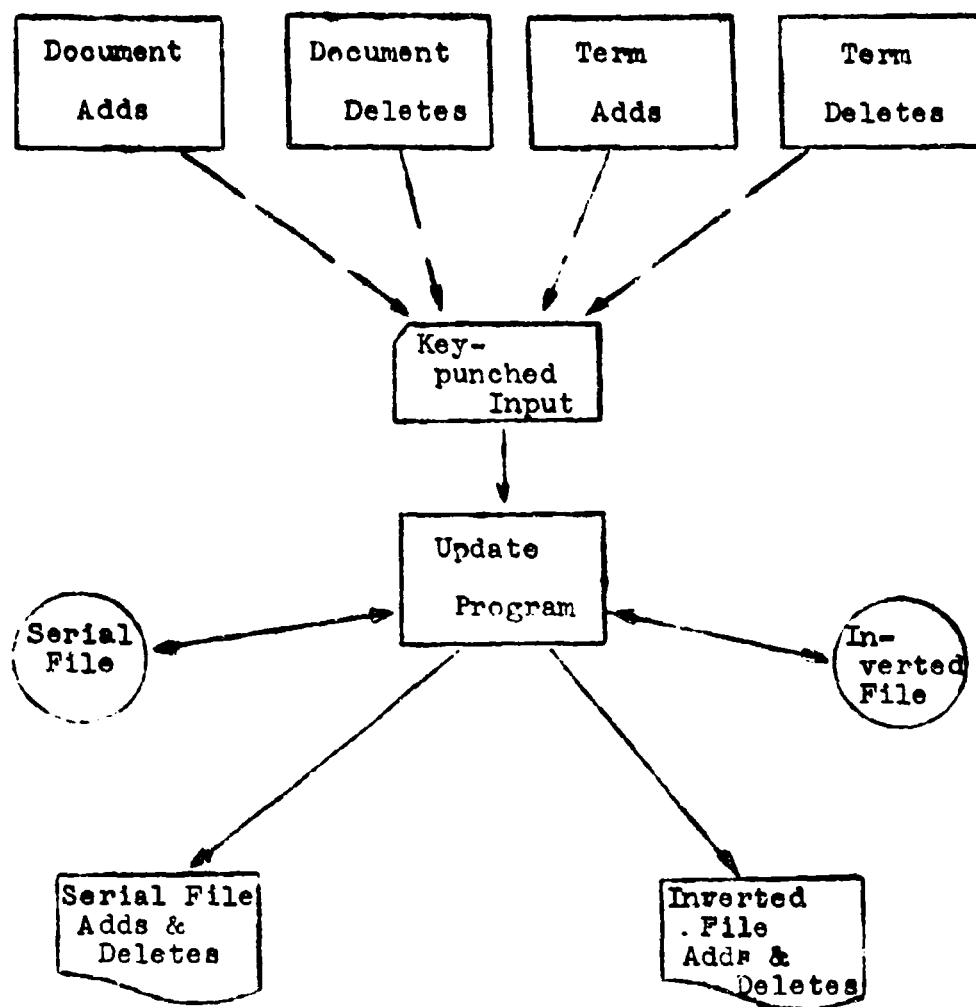
Figure 4. The Update Subsystem

Because the programs of updating are costly in time and money, batch-processing is necessary.

The Listing Subsystem

A series of listing programs are included to produce a book catalog of the documents in the system. Additional lists of the files are used as a verification to check update operations.

The output of the listing programs is from a high speed printer (900 lines/minute). An illustration of this subsystem is Figure 5.

B. Batch Processing Limitations

A number of general problems seem inherent in batch-processing systems. Some of the problems enumerated here are based on systems work connected with the Lehigh system and with other systems of a similar nature.

Flexibility: Little flexibility is available in computer processing. Once a single request or a batch of requests are submitted, the "die is cast". The user has no control over the processing of his request. The programming associated with tape systems is generally non-modular in structure. Modification of the system programs can generate lengthy changes which sometimes result in program redesign.[5]

Monitoring: Program execution cannot be monitored.[6] Programming work (debugging) in many cases becomes longer than the initial design and coding of the respective program. File updating is not guaranteed unless a listing of the file is generated after a batched update run.

Interpretation: The user must translate a request into the language or system code of a fixed format structure. Where requests may be complex, a human intermediary is consulted to help the user in transforming his request. In general, the more intermediaries interposed between the user and the
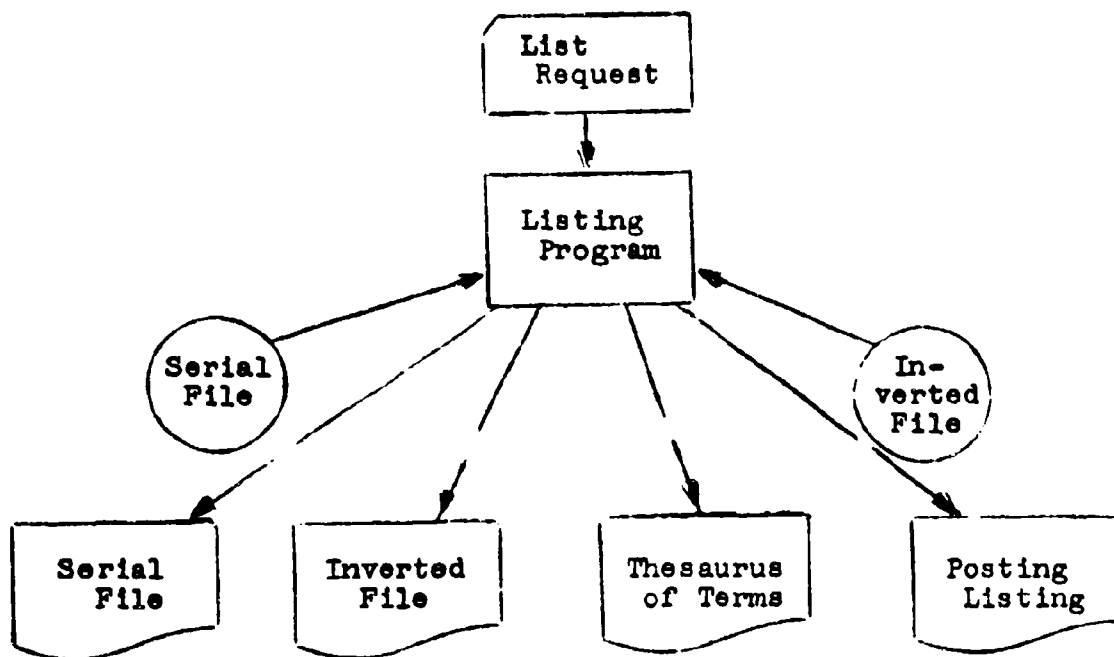
Figure 5.   The Listing Subsystem

document reference system, the greater the chance
for the distortion of the user's original intention
or request.[7]

Turnaround Time: This varies for a particular system,
but normally this time is not a positive attribute
of a batch processing reference system. The user
can become discouraged if his request is placed
in a long queue for processing. For the Center
for Information Sciences BPTO system, turnaround
time varies from one hour to three hours. This
is low compared to other batch-processing systems.[8]

In summarizing the problems connected with batch-
processing retrieval systems, Vincent Giuliano states:

> Most existing computerized document retrieval
> systems are run on a "closed shop" basis, and
> there is little or no facility for meaningful
> interaction between requestor and machine.
> The requestor usually specifies his inquiry by
> writing the key terms on a special form, which
> is later keypunched onto punched cards and
> verified. Later, still, the cards for several
> inquiries are batched, and the retrieval is
> accomplished by sequential computer searching
> of magnetic tape files. The answers are
> recorded on magnetic tape which is not printed
> until after the searching run, and the answers
> are finally mailed to the requestor. Although
> the computer may work in nanoseconds, typical
> cycle time from formulation of inquiry to
> receipt of answers may be anything from several
> hours to several days.[9]

C. The On-Line Concept

On-line computing is a concept that has existed since a printer, magnetic tape unit, or other peripheral device was connected directly to a computer. Only within the last few years has this term been used to denote a computer system that employs remote terminals or teletype printers. A general definition of on-line is as follows:

A system in which the input data enter the computer from their point of origin. The output data from the computer are transmitted directly back to the originating point.[10]

For the purposes of this paper, the definition will be modified to clarify the on-line concept in a reference system environment. The modified definition is:

A document reference system in which the input query enters the computer from a remote (different physical locale than the computer) terminal. The reply to the query is transmitted from the computer directly back to the same remote terminal.

Simply, the transmission controller in an on-line system is a switch which allows one or more remote terminals to be connected directly to a computer. When utilization of the computer from a remote terminal is requested, the controller after "polling" the other channels, selects the channel desiring service. The remote terminal is then on-line directly to the computer.

In this particular configuration, no other remote terminal can gain attention until service is completed at the other terminal. Figure 6 is an illustration of remote terminals on-line to a computer.

The remote terminal becomes both the input and output device in such a reactive system. Entry is made directly via the keyboard on the terminal. Output is also achieved through the relatively low speed printer which is adequate for reactive communication but not for volume printing.
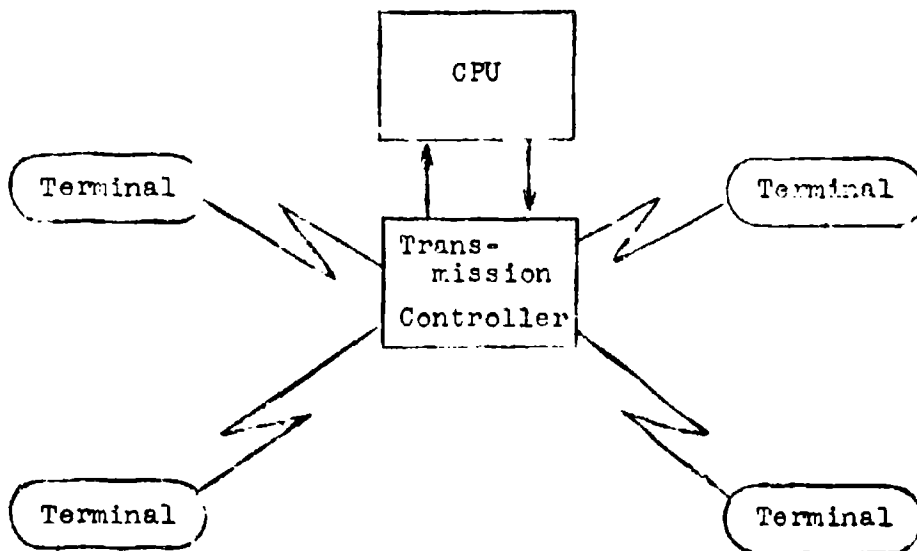
Figure 6. On-Line Remote Terminals

III. Transition to On-Line Retrieval[11]

A. Objectives of Conversion

In the following discussion, attention is focused on
the advantages gained in converting a document reference system
from BPTO to an OLRT system. These advantages are based on two
assumptions:

    (a) A document reference system is a _service_
       offered to its user community.

    (b) The service system should be as _user-oriented_
       as economically feasible.

Assumption (b) is a variable that can be determined by
analysis of an independent system environment. The user-oriented
concept of a service system has a _catalyst_, the OLRT concept.
The user is severely restricted in control of the information
retrieved in a batch-processing system environment. The system
procedures and formats the user must follow establishes the
retrieval interaction. Consequently, the control of the inform-
ation retrieved is a function of the system, _not_ the user.

The OLRT system returns control of the request process
to the user through a user-computer dialogue. The dialogue gives
the user the feeling that he controls the system. He is free
from details (procedures, formats, and codes), and more of his
attention can be concentrated on formulating search requests.

To achieve a better integration of user and system,
some general objectives must be established in the OLRT environ-
ment. These objectives must give the relationship a more mean-
ingful interaction, as well as lending flexibility to a document
reference system. Characteristics of this desired flexibility
are growth, flow of design, and experimentation.

Centralization

The objective in centralization is to place large
files on-line in order to give the system a common data base.

The centralization of files is limited by the <u>potential storage</u> of the system and, normally, has economic as well as physical boundaries.

Faster Response Time

It is obvious that for conversational mode, faster system response is necessary. This is a function of system design in both hardware and software configurations. The parameters involved in faster system response are file structure, search strategies, queuing algorithms, and hardware access times.

Integration of Functions

This objective is based on the processing operations of the present system. The batch-processed tape-oriented system has associated functions that must be performed by human intervention. The OLRT design (when properly designed) is a self-sustaining system whose various functions are integrated and dependent only on user manipulative operations. In contrast, the human operations or functions in a tape-oriented document reference system are:

1. Operation of computer and associated equipment;
    (a) initiating every system run;
    (b) changing tape reels;
    (c) distribution of output or run results;
2. Operation involving user-consulting (system custodian):
    (a) teaching new users;
    (b) helping user to interpret system codes, etc.;
    (c) phrasing user request;
    (d) submitting a user request, keypunching, etc.;

Along with the BPTO system functions, system failure should be included. If a failure is caused by equipment malfunction, the system is "down". No searches can be performed

until the failure is corrected. An OLRT system lends itself
to a degraded mode of operation. This is achieved through
fallback and automatic recovery procedures.[12] These procedures
allow the system to continue service, but, generally, result
in a slower response mode of operation.

System Accessibility at the User Location

The remote terminal brings the system to the user's
location. With remode location access to the system, user
communication with the system is of a higher degree than a
batch-processing system. Richard H. Wilcox states:

> Closely related to simple communications is
> physically convenient access. Experience with
> conventional computer facilities (batch-process-
> ing) indicates that their use varies roughly
> inversely with the distance from the user. [13]

System accessibility governs a system "usage function".
The resulting increased usage is a safe intuitive concept, but
system complexity must be considered. The system, depending on
its service function, can be designed for general user simplicity
or the design could be focused to serve a highly specialized
user community. Associated with these considerations is
whether the system teaches the on-line remote user or whether
a human intermediary is consulted on system usage.

B. Factors of Conversion

The factors of conversion are those major functions of
a total system that are affected by the conversion from a BPTO
system to an OLRT system. This section discusses the factors
that must be considered in converting a document reference and
retrieval system.

1) File structures
2) File addressing
3) Maintenance
4) Communication

5) Scheduling

6) Programming structures

The factors are discussed from the view of experience based on systems research work done at the Center for the Information Sciences. First, it is necessary to mention briefly some points about the hardware configuration of an OLRT system, for the hardware has a direct relationship to some of the factors of system conversion.

In order to make files accessible on-line to a search system, a random access disk storage unit is necessary. This type of unit is more efficient than tape drives due to a more direct means of file addressing. Another piece of hardware peculiar to an on-line configuration is the transmission controller. This unit is a computer or stored program multiplexer. Its purpose is to maintain communication control of remote terminals (i.e. terminal selection, code conversion, etc.) This multi-processing mode of operation frees the CPU of operations that require too much of its time. Based on the hardware requirements discussed above, Figure 7 illustrates an OLRT hardware configuration for a document reference service environment.

1. File Structures

The files and their structures are the foundation of any storage and retrieval system. These structures when properly designed yield efficient searching and storage operations for the system. They also give the system a degree of flexibility when supporting system manipulations that can add or delete file records without program modifications.

Using a disk unit for the storage of the files of the system, a different search strategy is used from the
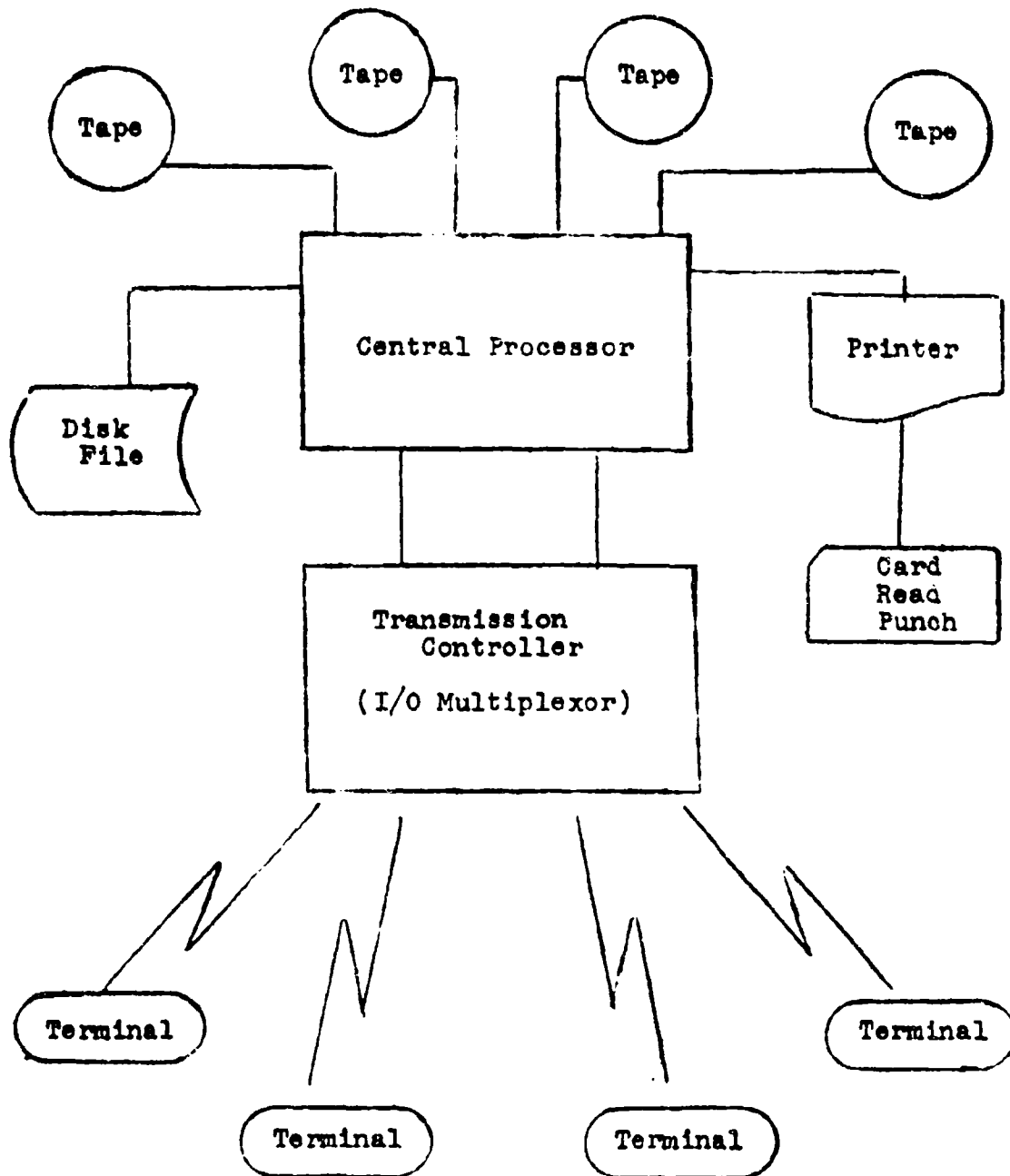
Figure 7. An OLRT Hardware Configuration

one discussed for the tape-oriented BPTO system. Disk search
is performed on an inverted file structure while serial files
are used subsequantly only for retrieving author, title, and
bibliographic data.

An algorithm is used to determine a disk address
for each particular document reference record identified by a
term in the search inquiry. The document reference record is
in a directory file whose structure can be or is the same as
the serial file. This inverted file search strategy is a more
direct means of accessing records on a disk which eliminates
searching of a complete file. The access or addressing
techniques for this type of system are discussed in the section
on File Addressing (below).

Although the BPTO system operates primarily on a
serial file, maintaining an inverted file is valuable if the
possibility exists for future system conversion. If en
inverted file is not properly maintained, it must be "created"
from the existing serial file at time of conversion.

Inverted File Structure

The inverted file structure of the BPTO system for
the most part remains the same when converted to disk. In
order to maintain storage efficiency on a disk, transformations
are used to "tie" disk records together by the technique of
chaining. Utilizing a chain approach, the inverted file
records do not have to be stored on the disk in sequence.
One approach in implementing this technique is to allow space
in the inverted file structure for an indicator. This
indicator is the disk storage location of the next or continued
record for that particular term. If no continued record
exists beyond the present record, a "flag", or termination

signal, is placed in the storage location.

If the physical disk record consists of a term number beyond the record size of sixty-four computer words, one word is the transformation or chain indicator and the other sixty-three words are document numbers. The last logical record containing document numbers has the stop "flag".

Serial File

The structure of the BPTO serial file can remain the same on disk for all practical purposes of _direct_ conversion. If Boolean search operations are performed on the inverted file, the term numbers for each document do not have to be included. The file/document can be compacted with other bibliographic data or additional information about the document.

However, if associative searching is used in the system, compacting must be disregarded due to the nature of associative searching and the need to determine co-occurrence of terms for each retrieved document reference.

2. File Addressing

The previous discussion on file structures in an OLRT system described the use of an inverted file structure for searching. Therefore, in the conversion of the BPTO system, the search statement remains the same (Boolean operators) but the addressing techniques used to retrieve the file data are different.

In the tape-oriented system, locating was the process of looking at each document in the file serially. Any computer memory location or disk location can be addressed by location. A tape can be addressed by searching for a number representing a type of record or block identification.

The scanning or searching of a tape serial file is
not a direct concept of addressing. When the files are
stored on disk in an OLRT system, the addressing techniques
are accomplished by five storage assignment techniques:[14]

       (a) Direct addressing

       (b) Algorithm calculation

       (c) Scanning

       (d) Table look-up

       (e) Randomizing

Most systems will use a combination of these address-
ing methods. Direct addressing is the easiest to implement and
most often is most economical. Utilizing the table look-up
technique, the computer searches the table rather than the
file.[15]

The other three techniques were impractical or pro-
mised to be time-consuming. For example, the algorithm is a
mathematical relation that generates a record location on
disk from an input reference. In the systems under discussion
here, the input reference is a natural language term or terms.
The algorithm could not differentiate between two terms that
are the same but of different meaning, such as _manual_ (by
hand) and _manual_ (a guide). Another complexity in formulating
an algorithm for addressing are when terms of different
spelling but containing the same letters (i.e. a permutation)
exist. The procedures necessary to achieve independence
in an algorithm become too great for practical and efficient
implementation.

Scanning or searching the files is a lengthy process.
The scanning method is not efficient when a great number of
files are maintained. The randomizing technique of addressing

has the disadvantage that the correct record will probably not be found the first time when addressing the file because of likely duplication in arrangement.[16]

The two methods, direct addressing and table look-up, when used together, become what shall be here called a "storage map" for the inverted file. This approach yields an efficient as well as a flexible technique for retrieval of documents.

The Storage Map

The storage map is analogous to the index or table of contents of a book. It conveys to a searching operation where a particular record is located on the disk. For a coordinate indexing system, the index terms and their respective disk locations appear in a table structure (Figure 8, below).

Terms	Locations
Abstract	062001
Computer	039201
Program	053920
Systems	056210
Xerox	062901
.	.
.	.
.	.

Figure 8. Storage Map Organization

When the Boolean search statement enters the system,
it is inspected for the operators. Once the operators for the
request are determined, the storage map is linearly searched
for the respective index terms of the statement. When a
match occurs, the operation becomes one of direct addressing
the location on disk of the first record for the index term.

If file updating is done on magnetic tape, tape-to-
disk routines must be employed. The tape updating may affect
the storage map disk locations. Each update of the files
must be checked against the storage map to see if a term
record has been added or deleted. The locations of the term
records on disk do not have to be sequential but when file
updating on magnetic tape is used, sequential locations are
advisable.

This technique of using a storage map gives a system
an efficient and direct method of accessing disk records.
The elimination of searching files yields faster response
times to a request. Other information which might be stored
in this table rather than the file itself could be relevant
during retrieval procedures, such as:

> (a) Scope notes: Aids or additional clarifi-
>   cation of the index terms. By placing them
>   in the table, additional searching is not
>   necessary for a listing of the system's terms.
>
> (b) Document counter: This is used in determining
>   associativity relationships.
>
> (c) Last location: For updating purposes (new
>   terms, etc.), it is useful to know storage
>   restrictions for an efficiently packed disk.

3. Maintenance

A dynamic document reference collection requires that techniques for file maintenance must be included. The maintenance function for file updates employs basically the same logic as in the BPTO system. Procedures for updating change when updating is done from a remote console. An OLRT system whose expected service "never fails" must incorporate also a new complete set of special purpose utility programs such as diagnostics and automatic recovery programs. Most of these were not fully implemented because of the experimental nature of the system.

4. Communication

When a document reference system is converted to an OLRT mode of operation with remote terminals, one of the objectives is to establish as natural a man-system interface as possible. For this, a natural language vocabulary is needed. Many system operations must be put in the form of decisions or questions for the user. A conversation made up of questions by the computer and answers by the user is the result.

With real-time or conversational response environment, the user is given the opportunity to change his mind, to think of new requests, and literally to control the system. The ease of use combined with more system control give greater satisfaction to the user.

For the BPTO system discussed in this study, the request or input from the user is a Boolean polynomial search request utilizing term numbers. If the user can submit a request with natural language terms (e.g. automatic + indexing - abstracting), system utilization for the user can be made easier. The user in this latter case would not have to

interpret his statement in specialized term code numbers.
Extending this notion of natural language requests, a natural
language request sentence may be allowed (e.g. "I want all
documents indexed by automatic and indexing but not abstract-
ing."). This kind of request statement requires syntactical
analysis procedures to interpret the index terms as well as
the Boolean operat ·s.

With experience in establishing an on-line teaching
system for the user, a query-respond environment may be allowed.
This environment would force the user to give additional infor-
mation that would lead the system to a more relevent answer.
To achieve on-line user communication with an operating system,
an increase of programming logic and programming instructions
is required. These increases are determined by the programming
language used.

For the BPTO system, the on-line implementaticn of the
search programs (associative and non-associative) in the Lehigh
system resulted in an increase of one third in programming
statements. The BPTO search program contains 198 FORTRAN II
instruction statements. When the programs were converted to an
on-line conversational-mode with user control options (choice
of searches, output wanted, etc.), programming statements
numbered 296. This allowed the user to respond "yes" or "no"
to prompting questions.

5. Scheduling

Although scheduling is a programming function, it must
first be considered independently of programming structures.
Scheduling and the decisions associated with algorithms of
scheduling are matters of policy and form the core of any OLRT
system operation. The formulation of decision rules for

handling the system when multiple-users or terminals are
introduced must be evaluated and worked out before programming
the functions. The major considerations of scheduling are:

      (a) User priorities

          (1) external

          (2) internal

      (b) Interrupts

      (c) Queues

      (d) Main scheduling

      (e) Input/output scheduling

These are all issues associated with the operating
system of an on-line configuration. Conversion must provide
for control of these matters which were of far less importance
in a batch processing mode.

6. Program Structures

One of the major reasons that batch-processing
programs are not flexible in their structure is that they are
often not segmented into their logical parts. A batch-
processing program has a flow of procedures and operations
that terminate in some final result. A program written in a
straight flow mode executes its instructions to accomplish
one complete task. That is the purpose of the program. When
converting to an OLRT system, the programs have to be structured
for flexibility and modification.

Modular Structures

Modular structuring of programming procedures and/or
instructions divide the total program into its logical parts.
A logical part is a group of algorithms or instructions that
perform one or more functions on the data or program input.
These logical parts that combine to form a total program are
called modules. If for some reason a module or logical part

is revised or eliminated, other modules of the total program should not be affected. In an operational on-line system these modules are called and linked at compile and/or execute time to perform tasks within the system.

In a dynamic environment, a system must adapt to change. For example, a system may normally experience changes due to experimentation, policy changes, or new design strategies. The system designers, to meet this changing environment, must create programs that can be quickly understood, easily modified to reflect changing conditions, and whose effect on related programs is understood.

The construction of a modular main program in an OLRT system uses an "overlay" technique and provides an efficient way for programming computer core memory programs. It conserves memory by the use of a physical part of memory for more than one modular routine.

C. Programming Considerations

Although an on-line interactive mode usually requires additional equipment, programming is the main cost of system conversion.[17] To implement the BPTO system to an OLRT system is <u>not</u> program modification; it is programming systems redesign. The programming system's logic changes when flexibility and expanded service for a user-system interface are implemented for the OLRT environment. Five points of OLRT programming must be considered.

> (1) Computer manufacturers offer standardized system and application programs with computing and communication. These programs must be modified to implement a particular system.

(2) The interaction between user and the system
must be supported by messages to guide,
prompt or advise the user in his own
language about progress of the search.

(3) Real-time programming must have interrupts,
queuing, scheduling and, diagnostic techniques.

(4) Programs must become modular. System
variables can not be fixed in a dynamic
environment.

(5) Hardware compatibility is sometimes hard to
obtain. For example, if a Model 33 teletype
is used as a terminal, its codes must be
translated to the code scheme for the
particular computer. This requires a sub-
routine or modular routine of a table look-
up nature.

Before programming costs can be estimated, an eval-
uation of the programming personnel available must be reviewed.
Programmers can make a system conversion costly if their
work is not evaluated first. Five points can help to
evaluate present or future manpower output.[18] These are:

(1) Experience in programming;

(2) Efficiency in programming;

(3) Creativity;

(4) Degree of training needed;

(5) Compiler languages or machine languages used.

On the basis of these points, a programming output for a staff
may be estimated.

In the conversion of the CIS Document Reference
Retrieval System, emphasis was placed upon developing the
communication with the user in the sea--h routines.
Additional initial effort was placed in converting both the
serial and inverted fi's from magnetic tape to disk form.
Support programs to l st the Thesaurus from the inverted disk
file and new author listing program. During the period
June 1966 to April 1967, three graduate research assistants
instrumental in development of the batch processing system
spent about half-time each in developing and proving these
programs and in relating them to other aspects planned for the
OLRT capabilities.

Features included in the search program were:

(1) Inclusion of an associativity table print-
out option.

(2) Division of the program into overlay seg-
ments for integration into an anticipated
Datanet 15 Communication Subcommittee for
the GE 225.

(3) Reduction of line feed time on teletype
printout.

(4) Feedback of the equivalent search terms
for confirmation by the user.

Experience with this program indicates the following advant-
ages of the disk search over tape search:

(1) More natural and flexible man-system inter-
action; aided by entering only one search
at a time.

(2) Reduction of system time spent in actual
searching of the files by proportions up
to 100 to 1.

(3) Greater control over the validity of
search by immediate error checks and feed-
back messages.

(4) Convenient record and verification of the
actual descriptors represented by the term
numbers entered.

(5) The number or absence of documents yielded
by a search is known before time is spent
in the listing of results phase.

(6) A choice of three options for output format
and detail is offered.

(7) Printing of superfluous blanks in document
records has been eliminated in output.

(8) Modification and reformulation of the
search request based upon initial results
can be done immediately.

Utility programs to achieve file maintenance and up-
date through remote access were designed during the period
May-October 1967 in order to:

(1) Provide means for convenient, efficient,
and economic display and/or modification of
the data base file from on-line terminals.

(2) Provide a complete audit trail of the
additions, deletions and alterations made
to the files.

(3) Incorporate flexibility and modularity
to facilitate

(a) changes in record format and
file organization

(b) addition of new functions

(c) independent programming and

testing of software modules.

The plan called for functions in the initial phase to add, delete, edit and display individual document and term records and to add or delete indexing relationships.

Estimates of the programming effort to develop these data management functions and to revise the GE 225 monitor to include the Datanet 15 far exceeded the amount of personnel time available for the remainder of the project. These features were basic to the operational aspects of maintaining the system and would have less effect upon the search user.

The system was experimental and was to be phased into other work and the GE 225 was to be replaced. Therefore, these latter programs were conceived but not implemented.

IV. User Reaction

A. Study Procedure

To assess the user reaction to the on-line system, a structured questionnaire was administered to a selected group of subjects after using the Center for the Information Sciences system. The questionnaire asked the subjects to evaluate the system performance on the basis of three major operating level components:

(1) System search procedures - vocabulary control, search logic, conversation routine, document collection, indexing, and associativity.

(2) Hardware facilities - terminal and communication line.

(3) Software capabilities - access to system, response time, and program performance.

The reply options available to the subjects were: favorably impressed, satisfactory, needs improvement, severe problem, and no opinion. In addition, ample space was provided for comments. A supplement was issued with the questionnaire presenting suggestions to the subjects for the purpose of assisting in the formulation of their evaluation and generating comments.

To establish a natural frame of reference, the questionnaire was structured according to the standard search procedure by which a person having an actual information need would use the system. The procedures included were:

(1) Preparation of request.

(2) Submission of request.

(3) Initial search.

(4) Retrieval.

(5) Reformulation of request.

(6) Termination of search.

The subjects rated the components of the system for each of these procedures, although not every component was applicable to each procedure. A sample questionnaire appears as Appendix II.

In order to assist evaluation, the study was conducted by grouping the participants who responded to the questionnaire. Each subject was given a retrieval question representing an information need. Using a list of the index terms available in the system and a listing of the contents of the inverted file, the subjects were asked to formulate a search request for this information need which would retrieve between five and twenty documents. Once the subject had formulated his request, he was given an opportunity to query the CIS collection using the on-line system, learn the number of documents identified and, if necessary, reformulate the original request until he satisfied the conditions of the experiment. No assessment of document relevance was made; instead, the subjects were told to assume the system was operational and they were approaching it as though the query represented their own requirement for information.

Replies to the questionnaires were scored on the basis of the following scale:

5 = favorably impressed
4 = satisfactory
3 = no opinion
2 = needs improvement
1 = severe problem

Each subject's evaluation of the CIS on-line system was calculated from this assignment of values and is presented as Appendix III.

Twelve subjects were divided equally into four groups, and separate queries were assigned once per group. The primary purpose of the study was to expose the subject in depth to the features of the system. Therefore, the queries were purposely selected such that there was a high probability that a subject would have to reformulate his search request in order to retrieve between five and twenty documents. This also increased the likelihood that the associativity search procedure would be used. Of

secondary interest in the study were the times required
to formulate the search request and to complete success-
fully the search requirements.

In view of the fact that all of the subjects were
not acquainted with document reference and retrieval systems,
a brief period of orientation preceded each experimental
session. An oral presentation, approximately forty-five
minutes in length, was given which outlined the concepts and
techniques employed in the CIS system. This was
accompanied by an on-line demonstration of the capabilities
of the system. In addition, an introductory supplement,
detailing the functions of document reference and retrieval
systems in general, and the CIS system in particular, was
distributed one week before the experimental session was
conducted.

B. Selection of Subjects

Two factors were considered in the selection of
the twelve participants: (1) the subject's knowledge of
document reference and retrieval systems, and (2) the
subject's experience with on-line systems. Two levels of
each factor were used, making a total of four unique groups
with three subjects in each group. The characteristics of
each group are shown in Figure 9.

Knowledge of Document Retrieval Systems

		High	Low
On-Line System Experience	High	I	II
	Low	III	IV

Figure 9. Experimental Subject Groups

Subjects from Groups I and III were selected from
eligible graduate students in the Information Sciences
program who had not been involved with the design and imple-
mentation of the CIS system. In all cases, the on-line
experience of Group I subjects was obtained from using the
CIS system in addition to other on-line information systems.
Group III members had seen the CIS on-line system
demonstrated, but had never used it themselves. Group II
and IV participants were selected from volunteers in a
graduate level course in industrial information systems.
The on-line experience of Group II members had been obtained
primarily from contact with time sharing and communication
systems.

C. Results

For purposes of analysis, evaluations from the
appropriate groups were combined according to the experience
factors under consideration, and are presented below in
tabular form. Table I compares the levels of experience
with on-line systems and Table II shows the results obtained
when the two levels of knowledge of document reference and
retrieval systems are compared. Evaluations for each of the
six phases of the search procedure are presented, together
with an average for the entire system.

Groups	Prepare Request	Submit Request	Initial Search	Re-trieval	Reform Request	Terminate Search	Avg.
I and II	3.39	4.38	4.4	3.85	4.0	4.22	4.04
III and IV	3.67	3.71	3.95	3.C	3.85	3.89	3.81

Table I. Experience with On-Line Systems
(Groups I and II High)

Groups	Prepare Request	Submit Request	Initial Search	Re-trieval	Reform Request	End Search	Avg.
I and III	3.17	3.92	4.03	3.73	3.81	4.11	3.81
II and IV	3.89	4.17	4.27	3.92	4.04	4.0	4.04

Table II.   Knowledge of Document Reference
and Retrieval Systems
(Groups I and III High)

The average times required for request formulation
and the on-line search procedure during the experiment are
grouped according to the two factors under consideration,
and are shown in Table III.   A detailed breakdown of these
times for each of the twelve subjects is given as Appendix
III.

Experience		Groups	Request Formulation	On-Line Search
On-line	High	I and II	9:33	12:10
	Low	III and IV	14:22	14:05
Document	High	I and III	7:38	13:55
Retrieval	Low	II and IV	16:18	12:20

Table III.   Times for Request
Formulation and On-Line Search
(Minutes)

Other experimental results of interest, include
the actual requests formulated by the subjects and the
comments given in response to the questionnaire. Both of
these items were considered to be potentially significant
and were used to supplement the interpretation of the
experimental results.

D. Discussion of Results

Preparation of Request

It is apparent from the results that all groups were the least satisified with this phase of the CIS on-line system. The most notable example of this comes from Table II, where a higher level of knowledge of document reference and retrieval systems resulted in by far the lowest rating of any phase of the search procedure. In contrast, the groups representing a lower level of experience rated the preparation of request phase much higher. One likely explanation is that knowledgeable people were more critical because they were aware of alternative methods which might prove more satisfactory than the existing method.

Although any results pertaining to times derived from this experiment not definitive, Table III provides some interesting figures in regard to request formulation. When Groups I and III are compared to Groups II and IV, it is noted that a previous knowledge of document reference and retrieval systems resulted in a reduction of approximately one-half the time required to formulate a request. This is probably most accurately attributed to a familiarity with the vocabulary used by the system.

In attempting to pinpoint the main areas of dis-content with this phase of the search procedure, the comments from the questionnaires are particularly interesting. The comments most frequently cited are the preference of natural language over numbers for index terms, the lack of a logical organization of the thesaurus, the inadequacy of the scope notes, and the fact that the current arrangements for using the system (scheduling in advance) would be unsatisfactory if the system were operational. As might be expected after examining the ratings in Tables I and II, the portion of the questionnaire dealing with request preparation was the recipient of the most critical comments.

It should be mentioned that, with few exceptions, all initial requests formulated were too specific for the conditions of the experiment. That is, the subjects retrieved less than five documents. The exceptions were all submitted by subjects from the higher level of knowledge of document reference and retrieval systems.

Submission of Request

For the two levels of experience with on-line systems, Table I shows that the largest difference in ratings occurred in this phase of the search procedure. In addition, only the groups having little or no previous experience with on-line systems rated this below the group average. These two facts could serve to indicate that the terminal is cumbersome to work with until the user becomes familiar with it.

This conjecture is borne out in part by the comments appearing on the questionnaires. The main complaints suggested that the terminal was confusing to operate and tha the routine to analyze the request was too sensitive to errors. Just as common were the favorable comments that the conversation routine made the system easy to use and that the program performance was exceptionally fast for the computer being used.

Initial Search

This phase of the search procedure received the highest rating in three of the four groupings shown in Tables I and II. A deeper analysis of the questionnaires received indicated that all three system components being judged, hardware, software, and the system itself, were evaluated about equally by the subjects.

Most of the comments centered on the associative search capabilities of the system. Although it was generally credited with being a good approach for learning about the

way index terms are used in the system, it was also noted
that an associativity table is of little value when it is
based on a small number of documents.

Retrieval

In general, there was little disagreement among the
various groups in their evaluation of this phase of the
search procedure. The majority of the comments referred to
the time-consuming print-out of the options available for
presentation. It was also commonly noted that, although the
document collection coverage was adequate to test the concept,
it was too limited for an operational environment.

Reformulation of Request

Surprisingly little difference in rankings occurred
in the evaluation of this phase of the search procedure.
Since a certain amount of thought is required to successfully
reformulate a request, this was interpreted as a healthy
indicator that a user is not hindered in the use of the on-
line system by his lack of experience.

The comments indicated that associativity was a
useful tool in the reformulation of the initial request, but
was not always adequate in locating alternative index terms.
Some confusion about the associativity of related terms was
mentioned.

Termination of Search

This phase of the search procedure was included in
the questionnaire primarily for the sake of completeness and
resulted in rankings near the group average and relatively
few comments. In fact, the only comment generated mentioned
that the conversation routine was easy to understand.

The majority of the annoyances indicated on the questionnaire dealt with the system itself and less so with the associated hardware and software components. It is significant to point out that most of these annoyances can be easily remedied in an operational system and the remainder of them tend to disappear as the familiarity of the user with the system increases. The major complaints, not unexpectedly, centered on the vocabulary control employed by the system-- the use of term numbers rather than natural language and the absence of a logical classification of the terms in the thesaurus. The need for a more sophisticated operating environment was singled out as a prerequisite for expansion of the system beyond the experimental stage. An unexpected result was the apparent heavy reliance upon scope notes in formulating a request. Similarly, the complaint that the numerical format for request formulation was too stringent was not anticipated.

V. Conclusions

Generally speaking, the user reaction to the CIS on-line document reference and retrieval system was favorable, based on the results of this study. Little difference was noted in the average rating given on the questionnaire, regardless of the user's knowledge of document reference and retrieval systems or his level of experience with on-line systems. All groups tested rated the system near the "satisfactory" range, and relatively few "severe problems" were mentioned. It would appear from these results that, from the standpoint of a user, the CIS system has been successful in providing man-machine interaction with the contents of a coordinate indexed document collection.

It is important to understand that system conversion from a batch-processing document reference system to an on-line mode requires a system redesign to maximize benefits. The magnitude of the conversion factors to achieve user-system interface and expanded service is too great for the considerations of a <u>direct</u> conversion of an existing batch-processing system. That is, moving from a tape-oriented to an on-line system requires too many basic design decisions to allow mere amplification of earlier system. The logic of the tape-oriented search software requires modification. New techniques of file searching and file maintenance must be developed. New programming systems of scheduling, maintenance diagnostics, and communication must be established.

Because conversion to on-line system is a redesign function, the costs for conversion are high in relation to design of the initial system. Over half of the costs are those for the designing and programming of the system's software. The level of these costs will depend on the

desired system sophistication. Communications oriented
terminals and computer hardware are necessary to support
the system.

The potential benefits of an interactive system with
associativity searches are substantial, in that access is
provided to a document collection without the aid of a
human intermediary. The direct access capability of disk
storage provides for immediate retrieval of document
representations in response to the request of the user.
Equally important, the storage requirements are not pro-
hibitive. The concept of associativity is valuable in
adapting to the needs of both the uninitiated and regular
user. It provides information about the system vocabulary
which can be used to formulate a request or to adjust the
volume of output generated by the initial request. Lastiy,
the techniques used in this system can be employed in any
application dealing with documents and the assignment of
descriptors to those documents through a coordinate index-
ing approach.

APPENDIX I

EXAMPLE OF SEARCH

An example of on-line inquiry and search teletype output using the CIS Document and Reference Retrieval System is covered on the next two pages. It illustrates features implemented in these studies and discussed in the text of this report. It can be contrasted with earlier examples in the Systems Manual of April 1967.

THIS IS THE CIS DOCUMENT REFERENCE SEARCH SYSTEM

   DO YOU WANT AN ASSOCIATIVE SEARCH
?YES

   PLEASE TYPE YOUR SEARCH STATEMENT.
?(52800V90700)+(72500V52200)

   THE TERMS REPRESENTED IN YOUR SEARCH STATEMENT ARE --

      52800 - AUTOMATA
      90700 - TURING MACHINE
      72500 - LIST PROCESSING
      52200 - ARTIFICIAL INTELLIGENCE

   ARE THESE TERMS O.K. WITH YOU
?YES

   YOUR REQUEST YIELDS   4 DOCUMENTS.

             T E R M   A S S O C I A T I O N S

TERM NO	TOT. OCC.	CO-OCC.	TERM	ASSOCIATIVITY
52200	54	4	ARTIFICIAL INTELLIGENCE	0.0741
52800	17	2	AUTOMATA	0.0588
64500	10	1	GAME	0.0250
71500	36	4	LEARNING	0.1111
72900	92	3	MACHINE	0.0245
88600	11	1	TEACHING	0.0227
89600	10	1	THINKING	0.0250
90700	7	2	TURING MACHINE	0.1429

   DO YOU WANT TO REFORMULATE YOUR CURRENT SEARCH STATEMENT
?NO

PLEASE SELECT ONE OF THE FOLLOWING OUTPUT OPTIONS.   INDICATE YOUR
CHOICE BY TYPING THE APPROPRIATE NUMBER.
      (1)   DOCUMENT ACCESSION NUMBER ONLY.
      (2)   DOCUMENT ACCESSION NUMBER, AUTHOR, AND TITLE.
      (3)   ALL OF THE ABOVE WITH BIBLIOGRAPHY.
?3

CIS DOC. NO. F   891
KOCHEN, M.   A MODEL FOR THE PROCESS OF LEARNING TO COMPREHEND
  IBM RES CENTER, YORKTOWN HEIGHTS, NY, AFCRL-64-87, AF 19(638)-2752
APR 1964

CIS DOC. NO. B 1230
TOU, J. AND R. WILCOX EDS. COMPUTER AND INFO SCIENCES
  SPARTAN BOOKS INC, WASHINGTON

CIS DOC. NO. B 1984
TURING, A. COMPUTING MACHINERY AND INTELLIGENCE
  IN FEIGENBAUM,E. AND J. FELDMAN,EDS, COMPUTERS AND THOUGHT,MCGRAW
HILL, 1-35

CIS DOC. NO. B 1985
NEWELL,A. SHAW,J. SIMON,H.   CHESS-PLAYING PROGRAMS AND THE PROBLEM OF
  COMPLEXITY
  IN FEIGENBAUM,E. AND J.FELDMAN, EDS, COMPUTERS AND THOUGHT, MCGRAW
HILL, 39-70

  DO YOU WANT TO ENTER ANY MORE SEARCH STATEMENTS
?NF

  GOOD-BYE.   COME BACK AGAIN.

APPENDIX II

Center for the Information Sciences
Lehigh University
Bethlehem, Pa. 18015

You are being asked to evaluate the CIS experimental document reference and retrieval system on the basis of the following components. Included below are suggestions to help formulate your evaluation or generate comments.

I. System

A. Vocabulary control

1. index terms in the thesaurus

2. use of term numbers for searching (rather than natural language)

3. index term qualifiers, or scope notes

4. coordinate indexing approach

B. Search logic

1. Boolean operators used (+, V, -)

2. four term limit

C. Conversation routine

1. order in which questions are presented

2. adequacy of questions

3. formats

D. Document collection

1. coverage

2. quality of documents retrieved

E. Indexing

1. consistency

2. too heavy or too light

F. Associativity

1. as tool for search reformulation

2. relatedness of query terms and other index terms (Stiles' formula)

3. quantity of terms displayed

II. Hardware

    A. Terminal

        1. display purposes

        2. keyboard for input

        3. speed

    B. Communication line

        1. time between placing initial call and beginning of program type-out

        2. hardware interruptions during search procedures

III. Software

    A. Access to system

        1. waiting time to use system

        2. turnaround time

    B. Response time

        1. time between entering request and learning number of documents which satisfy request

        2. time between learning number of documents which satisfy request and print-out of document representations

    C. Program performance

        1. failure to perform as designed

        2. design limitations

SEARCH PROCEDURES

	Favorably Impressed	Satisfactory	Needs Improvement	Severe Problem	No Opinion
I. Preparation of Request					
A. System component					
1. Vocabulary control	☐	☐	☐	☐	☐
2. Search logic	☐	☐	☐	☐	☐
Comments:					
B. Hardware (none)					
C. Software					
1. Access to system	☐	☐	☐	☐	☐
Comments:					
II. Submission of Request					
A. System					
1. Conversation Routine	☐	☐	☐	☐	☐
Comments:					
B. Hardware					
1. Terminal	☐	☐	☐	☐	☐
2. Communication line	☐	☐	☐	☐	☐
Comments:					

	Favorably Impressed	Satisfactory	Needs Improvement	Severe Problem	No Opinion
SEARCH PROCEDURES					
C. Software					
1. Program performance	☐	☐	☐	☐	☐
Comments:					
III. Initial Search					
A. System					
1. Conversation routine	☐	☐	☐	☐	☐
2. Associativity	☐	☐	☐	☐	☐
Comments:					
B. Hardware					
1. Communication line	☐	☐	☐	☐	☐
Comments:					
C. Software					
1. Response time	☐	☐	☐	☐	☐
2. Program performance	☐	☐	☐	☐	☐
Comments:					

SEARCH PROCEDURES

IV. Retrieval

	Favorably Impressed	Satisfactory	Needs Improvement	Severe Problem	No Opinion
A. System					
1. Conversation routine	☐	☐	☐	☐	☐
2. Document collection	☐	☐	☐	☐	☐
3. Indexing	☐	☐	☐	☐	☐
Comments:					
B. Hardware					
1. Terminal	☐	☐	☐	☐	☐
2. Communication line	☐	☐	☐	☐	☐
Comments:					
C. Software					
1. Access to system	☐	☐	☐	☐	☐
2. Response time	☐	☐	☐	☐	☐
3. Program performance	☐	☐	☐	☐	☐
Comments:					

SEARCH PROCEDURES

V. Reformulation of Request

	Favorably Impressed	Satisfactory	Needs Improvement	Severe Problem	No Opinion
A. System					
1. Vocabulary control	□	□	□	□	□
2. Search logic	□	□	□	□	□
3. Conversation routine	□	□	□	□	□
4. Associativity	□	□	□	□	□
Comments:					
B. Hardware					
1. Terminal	□	□	□	□	□
2. Communication line	□	□	□	□	□
Comments:					
C. Software					
1. Response time	□	□	□	□	□
2. Program performance	□	□	□	□	□
Comments:					

SEARCH PROCEDURES

VI. Termination of Search

	Favorably Impressed	Satisfactory	Needs Improvement	Severe Problem	No Opinion
A. System					
1. Conversation routine	☐	☐	☐	☐	☐
Comments:					
B. Hardware					
1. Communication lines	☐	☐	☐	☐	☐
Comments:					
C. Software					
1. Program performance	☐	☐	☐	☐	☐
Comment:					

APPENDIX III

USER REACTION SCALE

Group	Subject	Prepare Request	Submit Request	Initial Search	Re-trieval	Reform Request	Terminate Search	Avg.
I		2.78	4.25	4.27	3.83	3.83	4.22	3.86
	1	3.0	4.25	4.6	4.0	3.75	4.0	
	2	2.67	4.25	4.0	3.75	4.13	4.67	
	3	2.67	4.25	4.2	3.75	3.88	4.0	
II		4.0	4.5	4.27	3.88	4.17	4.22	4.14
	1	3.0	4.25	4.4	3.75	4.0	4.0	
	2	4.33	4.75	4.2	4.0	4.13	4.33	
	3	4.67	4.5	4.2	3.88	4.38	4.33	
III		3.55	3.58	3.67	3.63	3.79	4.0	3.70
	1	3.67	4.5	3.6	4.0	4.13	3.67	
	2	3.33	2.5	3.8	3.13	3.63	4.0	
	3	3.67	3.75	3.6	3.75	3.63	4.33	
IV		3.78	3.83	4.27	3.96	3.91	3.78	3.95
	1	3.67	4.0	4.4	4.38	4.25	4.0	
	2	4.0	3.75	4.2	4.0	3.88	4.0	
	3	3.67	3.75	4.2	3.5	3.63	3.33	

Rating Scale:

5 - Favorably impressed   4 - Satisfactory   2 - Needs improvement
3 - No opinion   1 - Severe problem

APPENDIX IV

USER TIMES
(Minutes)

		Request Formulation	Search Procedure	AVERAGE TIME Request Formulation	AVERAGE TIME Search Procedure
Group	Subject	Request Formulation	Search Procedure	Request Formulation	Search Procedure
I				6:41	12:03
	1	5:10	12:26		
	2	2:39	13:32		
	3	11:45	10:12		
II				12:25	12:17
	1	9:12	12:55		
	2	12:47	11:11		
	3	13:55	11:25		
III				8:34	15:47
	1	10:22	11:35		
	2	10:05	12:32		
	3	4:35	22:35		
IV				20:10	12:23
	1	18:10	13:30		
	2	18:45	7:25		
	3	22:14	14:55		

REFERENCES

1. Taylor, Robert S., "System's Manual: Center for the Information Sciences." Unpublished research system manual for the document retrieval system, Center for the Information Sciences, Lehigh University, Bethlehem, Pa. 1966.

2. Jennings, Michael A., "Construction of Boolean Search Operations in a Coordinate Indexing System," in American Documentation Institute, Progress in Information Science and Technology, Proceedings of the 1966 Annual Meeting, October 3-7, 1966, Adrianne Press, 1966, pp. 17-27.

3. Curtice, Robert M., "Magnetic Tape and Disc File Organizations for Retrieval." Report No. ..., Experimental Retrieval Systems Studies, Center for the Information Sciences, Lehigh University, Bethlehem, Pa., July 1966.

4. Anderson, Ronald R., "An Associativity Technique for Automatically Optimizing Retrieval Results." Report No. 3., Experimental Retrieval Systems Studies, Center for the Information Sciences, Lehigh University, Bethlehem, Pa., April 1967.

5. Drew, D. L., Summit, R. K., Tanaka, R. I., and Whiteley, R. B., "An On-Line Technical Library Reference Retrieval System," American Documentation, Vol. 17 (1), 3-7 (January 1966).

6. Ibid., p. 3.

7. Ibid., p. 3.

8. Berul, Lawrence, Information Storage and Retrieval: A State-of-the-Art Report, Philadelphia, Pennsylvania: Auerbach Corporation, 1964.

9. Giuliano, V. E., "Automatic Message Retrieval by Associative Techniques," in Joint Man-Computer Languages, Massachusetts: Mitre SS-10, 1962, p. 5.

10. Martin, James, Programming Real-Time Computer Systems, New Jersey: Prentice-Hall, Inc. 1965, p. 4.

11. Jennings, Michael A., "An Analysis of the Factors of Conversion from a Batch Processing, Tape Oriented Retrieval System to an On-Line, Real Time System." M. S. Thesis, Center for the Information Sciences, Lehigh University, Bethlehem, Pa., June, 1967.

12. Martin, James, op. cit., p. 131.

13. Wilcox, Richard H., "Computer Augmentation of Human Reasoning," in A. Kent and O. Taulbee, Editors, Electronic Information Handling, Washington, D. C.: Spartan Books, Inc., 1965. p. 267-270.

14. Martin, James., op. cit., p. 100-103.

15. Martin, James, op. cit., p. 105.

16. Martin, James, op. cit., p. 107.

17. Head, Robert V., Real-Time Business Systems, New York: Holt, Rinehart, and Winston, Inc., 1965, p. 159-161.

18. Martin, James, op. cit., p. 349.

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)	2a. REPORT SECURITY CLASSIFICATION
Center for the Information Sciences Lehigh University Bethlehem, Pennsylvania 18015	UNCLASSIFIED
	2b. GROUP

3. REPORT TITLE

TRANSITION TO ON-LINE INFORMATION AND REFERENCE RETRIEVAL

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Final Summary, 1 June 1966 - 31 May 1968

5. AUTHOR(S) (First name, middle initial, last name)

William A. Smith, Jr., Ronald R. Anderson, Michael A. Jennings

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
26 August 1968	62 pages	

8a. CONTRACT OR GRANT NO	9a. ORIGINATOR'S REPORT NUMBER(S)
Nonr 610(08)	Experimental Retrieval Systems
b. PROJECT NO.	Studies, Report No. 4
c.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
d.	

10. DISTRIBUTION STATEMENT

Qualified requestors may obtain copies of this report from DDC.
No limitation on dissemination.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Information Systems Branch Office of Naval Research Washington, D.C. 20360

13. ABSTRACT

An existing experimental document and reference retrieval system operating in batch processing, tape oriented mode was converted to an on-line mode with user interaction. Objectives were faster response time, integration of functions, and system accessibility at user location. Effects on file organization, addressing techniques, file maintenance, communication, scheduling and programming are analyzed. Modification of an existing batch oriented system can yield an effective interactive search capability. Realization of full benefits of the on-line concept, particularly maintenance of the document base, requires system redesign rather than extension of existing procedures.

User reaction to the experimental system was favorable although a variety of improvements were suggested for an operational environment. Output of associativity indicators was valuable in reformulating search requests or in adjusting the volume of cited references for both experienced and unitiated users.

DD FORM 1 NOV 65 **1473**

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Document Retrieval						
Information Retrieval						
Information Science						
Search						
On-line Retrieval						
Associativity						
Reference Retrieval System						
Mechanized Retrieval						
Conversion of System						
User Reaction						
Interactive Retrieval						